A Problem for Guidance Control

A Problem for Guidance Control

Penultimate Draft
(final version in *The Philosophical Quarterly*, Vol. 58, No. 233, (October 2008): 685-692)

BY PATRICK TODD AND NEAL A. TOGNAZZINI

In this paper, we raise a worry for the influential theory of moral responsibility put forth by John Martin Fischer and Mark Ravizza (hereafter, F&R). In particular, we argue for the possibility of a case in which an agent is intuitively not morally responsible despite satisfying the conditions for guidance control, which F&R claim is the sort of control central to moral responsibility. After presenting the counterexample, we consider various ways in which the account of guidance control may be mended.

I. THE ELEMENTS OF GUIDANCE CONTROL

Before we can present the worry, however, we need to explain the basic structure and some of the details of F&R's theory.[1] F&R identify guidance control as the "freedom-relevant" condition on moral responsibility, and they spell it out as follows:

> (GC) An agent exercises *guidance control* of an action if and only if that action issues from the agent's own, moderately reasons-responsive mechanism.

The two crucial components of guidance control are *ownership* and *responsiveness to reasons*. But since our worry does not concern ownership, we need not discuss it here. We will focus instead on F&R's analysis of the sort of reasons-responsiveness they think is required for moral responsibility:

> (MRR) An actually operative kind of mechanism is *moderately reasons-responsive* if and only if: (1) it is at least regularly receptive to reasons, some of which are moral reasons, and (2) it is at least weakly reactive to reasons (but not necessarily moral reasons).[2]

Our worry targets F&R's account of (MRR). In particular, we will present a case according to which some agent's action issues from a mechanism that satisfies both (1) and (2) of (MRR), although we would intuitively judge the agent not to be morally responsible for the action. To understand our objection, then, it will help to dig a little deeper into F&R's theory by looking closely at the distinction between *receptivity* and *reactivity* to reasons.

Intuitively, to be receptive to reasons is to be able to recognize that certain facts about a particular situation are considerations either in favor or against taking a certain course of action. There are different degrees of receptivity to reasons, however. How should we specify the degree required for moral responsibility? F&R specify it as follows:

> (RRec) An actually operative kind of mechanism is *regularly receptive to reasons* if and only if:
>
> (1) There are possible scenarios in which (i) there is sufficient reason to do otherwise, (ii) the same kind of mechanism is operative, and (iii) the agent recognizes the sufficient reason to do otherwise, and
>
> (2) The possible scenarios described in (1) constitute an *understandable pattern* of reasons-recognition.[3]

Let us remark briefly on both components. Receptivity is a modal property of mechanisms. To find out whether a particular kind of mechanism has this property, component (1) tells us that there must be counterfactual circumstances in which, holding fixed the kind of mechanism, the agent in question manages to recognize the sufficient reasons to do otherwise in that circumstance. So, to take a concrete example, suppose that your actually operative mechanism issues in your purchasing a ticket to the Los Angeles Philharmonic for $50. In order for us to conclude that your actually operative kind of mechanism satisfies (RRec), one thing we need to know is whether something like the following counterfactual is true: if the ticket were to have cost $5 million instead (and the same kind of mechanism were to have operated), you would have recognized that this fact is a sufficient reason not to buy

the ticket. The truth of this counterfactual will give us a world (or sphere of worlds) at which you are presented with sufficient reason to do otherwise, and you succeed in recognizing it. Thus your actually operative kind of mechanism is, in a weak sense at least, receptive to reasons.

But more is required for (RRec) than one world at which you manage to recognize a sufficient reason to do otherwise. While this does amount to a *sort* of receptivity, it is not robust enough to ground moral responsibility. In particular, suppose that in addition to truth of the above counterfactual, the following counterfactual is true as well: if the ticket were to have cost $6 million instead (and the same kind of mechanism were to have operated), you would *not* have recognized that this fact is a sufficient reason not to buy the ticket. If we suppose that this counterfactual is also true, your mechanism begins to look quite mysterious indeed. Why would you recognize that $5 million is too much to pay for a ticket, but not recognize that $6 million is also too much? If you have no story to tell that would make this pattern of receptivity understandable, we should conclude that you are not in fact morally responsible for buying the ticket.[4]

In order to get around this problem, F&R add clause (2) to their account of reasons-receptivity. Clause (2) specifies the *regularity* component of (RRec): the situations in which you would and would not recognize sufficient reason to do otherwise must constitute an understandable pattern. There can't be any such oddities as recognizing that $5 million is too much, but not recognizing that $6 million is too much.

Let us now move to reasons-reactivity. An intuitive way to understand reactivity is, as F&R put it, in terms of a "capacity to translate reasons into choices (and then subsequent behavior)".[5] If I cannot translate reasons into action, then no matter how reasons-receptive my mechanism is, I cannot be appropriately held

morally responsible for what I do. We can spell out F&R's account of reasons-reactivity as follows:

> (WRea) An actually operative kind of mechanism is *weakly reactive to reasons* if and only if there is some possible scenario in which (1) there is sufficient reason to do otherwise, (2) the same kind of mechanism operates, (3) the agent recognizes the sufficient reason to do otherwise, and (4) the agent thus chooses and does otherwise for that reason.[6]

Clause (4) is crucial here. The first three clauses pick out one of the worlds that tell us about the mechanism's *receptivity*; if clause (4) is also satisfied at that world, then we know that the mechanism is appropriately *reactive*, as well.

What makes this an account of *weak* reactivity to reasons is the fact that only *one* possible scenario is needed in which clauses (1) through (4) are satisfied. (Contrast this with (RRec), according to which there must be a suitable *range* of worlds in which the agent recognizes sufficient reason to do otherwise.) The reason for this is that reactivity is "all of a piece" in the following sense: "If the mechanism can react to any reason to do otherwise, it can react to all such reasons."[7] This contention will prove important for the critique we present below.

F&R present a sophisticated and nuanced account of moral responsibility. Regularly reasons-receptivity and weak reasons-reactivity together constitute moderate reasons-responsiveness, which is one of the crucial components of guidance control, which itself is the freedom-relevant condition on moral responsibility. If an agent exercises guidance control of an action, and also satisfies the other conditions on moral responsibility (such as the epistemic condition), then the agent is morally responsible for that action.

II. A PROBLEM FOR GUIDANCE CONTROL

Now that we have sketched F&R's theory of moral responsibility, we can turn to our worry. Briefly, the worry is that the following sort of situation seems possible:

(1) An agent performs a morally reprehensible action despite the existence of reasons not to perform it,
(2) The kind of mechanism that actually issues in the reprehensible action is both regularly receptive to reasons and weakly reactive to reasons (as well as owned by the agent), and
(3) The agent *cannot* recognize the *actual* reasons not to perform the action.

If such a situation is possible, then F&R's account of guidance control is in trouble, because whereas (2) amounts to the claim that the agent exercises guidance control of his morally reprehensible action, (3) should lead us to judge that the agent is not morally responsible for his action. After all, if there was no way the agent could have recognized the reason to do otherwise that actually existed, how is it fair to blame him for what he did?

Let's attempt to be a bit more concrete, however, by instantiating the abstract structure given above. Imagine an agent – call him 'Tony' – who actually slaps you across the face. Now, as a matter of fact, Tony has reason not to do this. In particular, the fact that it will cause you pain is a reason not to slap you. But let's suppose that Tony cannot recognize that this fact is a reason not to slap you. That is, suppose that there is no possible world in which Tony's actual kind of mechanism operates, and in which Tony recognizes that the fact that his slapping you will cause you pain is a reason not to do it. Whatever reasons-responsiveness Tony might exhibit in other contexts, the fact that he cannot recognize your pain as a reason not to slap you should, intuitively, lead us to judge that Tony is not morally responsible for slapping you. (One might worry that this judgment is incorrect if Tony is nevertheless acting on a morally reprehensible reason. That is, one might think that acting on a morally reprehensible reason is enough to secure moral responsibility *despite* an inability to see the reasons to do otherwise.[8] Let us suppose, then, that when Tony slaps you, he's *not* acting on any morally reprehensible reason, but merely to amuse his friend.) However, and here's the important point, the mechanism that

actually issues in Tony's slapping you might nevertheless be moderately reasons-responsive. How?

Recall that (MRR) comprises two components: regular reasons-receptivity and weak reasons-reactivity. Let's take each in turn to see how Tony's mechanism might satisfy it. For Tony's mechanism to be regularly receptive to reasons, there must be possible circumstances, which together constitute an understandable pattern, in which there is some sufficient reason to do otherwise and in which Tony recognizes this reason. To see how this requirement might be satisfied, we just need to make Tony a bit weirder. For some inexplicable reason, you are special to Tony. Whereas Tony cannot recognize *your* pain as a reason not to slap you, he can perfectly well recognize Jessica's pain as a reason not to slap her, and Brian's pain as a reason not to slap him, and so on for every other agent. It's just that his recognitional capacities are for some reason blocked with respect to you. If that's right, then there will presumably be plenty of counterfactual circumstances in which Tony's mechanism is receptive to reasons to do otherwise. Moreover, there seems to be no reason to suppose that these circumstances do not constitute an understandable pattern. After all, even F&R point out that we can't expect *perfection* when it comes to receptivity. They say: "Of course, the regularity [of reasons-receptivity] need not be absolute; the mechanism must simply evince some suitable degree of regularity. Everyone makes some mistakes, and it is a matter of judgment precisely how much regularity is appropriate to require."[9] But if the regularity need not be absolute, then our case of Tony seems perfectly possible.

Reactivity is an even easier case. All we need is for there to be one possible circumstance in which the same kind of mechanism operates and in which Tony recognizes sufficient reason to do otherwise and succeeds in translating this reason into action. Nothing about Tony's local inability with respect to recognition appears

to affect his ability to translate reasons into action. Indeed, we can suppose that in at least one world in which he is confronted with the fact that slapping Jessica will cause her pain, not only does he recognize this as reason not to slap her, but he also refrains from slapping her for this reason. So it looks like Tony's mechanism may well be weakly reasons-reactive, as well.

But now we have arrived at a case that looks like a counterexample to F&R's account of guidance control. Tony acts from his own moderately reasons-responsive mechanism and yet, due to his local inability to recognize your pain as a reason not to slap you, we are inclined to judge that he is not morally responsible for slapping you.

We said above that F&R's claim about reactivity's being "all of a piece" is relevant to our criticism, and we are now in a position to see why. Since reactivity is all of a piece, all we need to know in order to conclude that Tony's mechanism is able to react to the actually present reason is that there is *some* world in which it reacts to *some* sufficient reason to do otherwise. It is thus impossible, on F&R's account, for there to be local deficiencies in reactivity. But they do not claim that *receptivity* is "all of a piece". In fact, they make the stronger claim that there is a crucial *asymmetry* between receptivity and reactivity on this very issue. It is this asymmetry that leaves room for local deficiencies in receptivity, and thus makes their account vulnerable to the sort of counterexample we have presented.

III. A POSSIBLE RESPONSE

It is worth considering one way that F&R may be tempted to respond to our counterexample. Recall that regular reasons-receptivity requires a range of possible scenarios that together constitute an *understandable pattern*. With this in mind, F&R might claim that their account *does* deliver the intuitively correct verdict that Tony is

not morally responsible since Tony's recognitional capacities are incoherent. As we have described the case, Tony suffers from a local inability with respect to recognition. While he can recognize that other people's pain is a reason not to slap them, he cannot recognize that *your* pain is a reason not to slap *you*. But this seems very odd. F&R might claim that any agent who exhibits this odd constellation of recognitional abilities does not act from a mechanism that is regularly reasons-receptive. In this way, they could deny that Tony is morally responsible.

The problem with this, however, is that it commits F&R to the claim that any situation in which that mechanism is operative is a situation in which Tony is not morally responsible for his action. This means that even in a world where Tony slaps *Jessica* (instead of you), he will not be morally responsible for doing so, despite the fact that he is able to recognize (and, we can suppose, in fact does recognize) that Jessica's pain is a reason not to slap her. This is also an intuitively incorrect result. The mere fact that Tony is unable to recognize your pain (and only yours) as a reason should not entail that he is not morally responsible in situations where he *does* recognize someone's pain as mattering.

In sum, the case of Tony presents F&R with the following dilemma. Either the pattern of Tony's receptivity is understandable, or it isn't. If it is, then according to F&R, Tony will be morally responsible for slapping you, despite the fact that he can't recognize that causing you pain is a reason not to slap you. If it isn't, then Tony won't be morally responsible for slapping Jessica, Brian, Sally, and so on, despite the fact that he can and does recognize that causing *them* pain is a reason not to slap *them*. Either way, F&R's account leads to an intuitively incorrect result. Hence, the case of Tony constitutes a counterexample to their account.

We do not think that our argument provides decisive reason to abandon their theory altogether, however, because there are a number of ways that F&R might

patch up their account. In the remainder of this paper, we will consider three possible fixes.

IV.PATCH #1: RECEPTIVITY IS "ALL OF A PIECE"

Perhaps the most straightforward way for F&R to avoid the counterexample we have presented would be simply to argue that both reactivity and receptivity are "all of a piece". Recall that our counterexample turned crucially on the following two claims: (1) There is a suitable range of possible scenarios in which Tony recognizes sufficient reason to do otherwise, and (2) Tony cannot recognize the *actual* sufficient reason to do otherwise. Making receptivity "all of a piece" would solve this problem, because the truth of the first claim would entail the falsity of the second.

This route has costs. First, F&R's "all of a piece" claim with respect to reactivity has been widely criticized by commentators, who will no doubt think that adding the same claim with respect to receptivity is equally unappealing.[10] Moreover, in recent work Fischer has argued that the fundamentals of their theory can survive jettisoning the controversial claim that reactivity is "all of a piece".[11] It would be undesirable, then, if F&R had to introduce the claim that receptivity is "all of a piece" in order to save their account of guidance control. Finally, even if one thinks that the claim about reactivity is plausible, the claim about receptivity seems less so. Do we really want to say that it's *impossible* for someone to have a patch of local blindness with respect to what reasons he can recognize?

V. PATCH #2: RECEPTIVITY IS "PART OF A PIECE"

Of course, instead of insisting on receptivity's being "all of a piece", F&R might instead say that receptivity is only "part of a piece". That is, they might claim that whereas it is not in general true that the ability to recognize some reason entails

the ability to recognize any reason, what is true is that when the reasons in question are of the same *type* (for instance, when they are about causing pain), the ability to recognize one reason of that type entails the ability to recognize any reason of that type. Recall that Tony recognizes that everyone's pain but yours is a reason not to slap him or her; he just doesn't recognize that *your* pain is a reason not to slap *you*. On this weaker "part of a piece" claim, F&R could maintain that this description of Tony is impossible, since it just doesn't make sense to suppose that someone could have a local inability in receptivity *within* a particular type of reason. Perhaps agents could be blind to whole classes of reasons, but within a particular class of reasons, it is impossible that an agent be able to recognize some members of that class but not others.

We think this "part of a piece" claim has considerable intuitive appeal. To see why, suppose that your friend breaks a promise to Jim and, when confronted about it, claims that he sees he shouldn't have done what he did because it constituted the breaking of a promise to Jim. But then suppose that your friend breaks a promise to Jill and, when confronted about *it*, claims that he can't see how his action's breaking a promise to Jill was a reason not to do it. The proper response to this friend, it seems, is to insist that either he *can* see the reason with respect to Jill (and he is just being obstinate) or that he was mistaken (or lying) about what he was able to see yesterday with respect to Jim. It seems implausible to suppose he sees it in one case but not in another that doesn't differ in any relevant way.

Unfortunately, however, adding this "part of a piece" claim to F&R's account fails to solve the fundamental problem, since we can simply reconstruct the counterexample at the level of *types* of reason.[12] That is, we can imagine a case in which Tony is blind to every instance of a particular type of reason (say, pain-based reasons) even though Tony is able to recognize instances of every other type of

reason. Now put Tony in a situation according to which the only reason to do otherwise is an instance of the type to which Tony is blind. If Tony acts contrary to this reason, he is intuitively not morally responsible for what he does, since he could not have recognized the reason in question. However, the mechanism on which he acts still seems suitably sensitive and understandable (and hence regularly receptive to reasons), since it is receptive to every other type of reason except the one in question, and (as F&R agree) we can't require perfection when it comes to receptivity.

VI. PATCH #3: RECEPTIVITY TO THE ACTUAL REASON

We will consider one final way to fix the account of guidance control. The real problem with the case of Tony is that he is unable to recognize the actually sufficient reason to do otherwise, and hence seems not to be morally responsible. An adequate account of reasons-receptivity, then, needs to require at least that the agent be able to recognize the *actual* sufficient reason to do otherwise. After all, when we are assessing an agent's moral responsibility, we should look at more than just whether the mechanism that issues in the action is functioning properly in some abstract sense; we should also look at whether the mechanism is appropriately connected to the *actual* circumstances. While (RRec) and (WRea) tell us when a mechanism is functioning properly in the abstract, they don't connect up to the actual circumstances.

So instead of revising (RRec), we need a *supplementary* property of the mechanism in question. It must be regularly receptive to reasons, weakly reactive to reasons, and, let's say, *receptive to the actual reason*:

> (RAR) An actually operative kind of mechanism is *receptive to the actual reason* if and only if there is a world in which the same mechanism operates and the same sufficient reason to do otherwise is present and the agent recognizes the sufficient reason to do otherwise.

If we add (RAR) to F&R's account of moderate reasons-responsiveness and guidance control, then the case of Tony is no longer troubling. Since his mechanism doesn't satisfy (RAR) with respect to his slapping you, he is not morally responsible for what he does.

VII. CONCLUSION

In this paper, we have presented a counterexample to F&R's account of guidance control that exploits a local blindness in reasons-receptivity. After presenting the counterexample, we considered three ways to mend the account of guidance control: claiming that receptivity is "all of a piece", claiming that receptivity is "part of a piece", and requiring receptivity to the actual reason. We conclude that the best way to fix F&R's account of guidance control is to supplement it by requiring receptivity to the actual reason.[13]

[1] For the theory in all its systematic glory, see John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, (Cambridge: Cambridge University Press, 1998). Our exposition of the theory draws directly from this book.

[2] Fischer and Ravizza, *Responsibility and Control*, pp. 69-85.

[3] Fischer and Ravizza, *Responsibility and Control*, pp. 69-73.

[4] You need not actually tell the story, but you need to be, in an appropriate sense, disposed to tell such a story. F&R think of this requirement in terms of a hypothetical interview *ex post facto*. See *Responsibility and Control*, p. 71.

[5] Fischer and Ravizza, *Responsibility and Control*, p. 69.

[6] Fischer and Ravizza, *Responsibility and Control*, pp. 73-76.

[7] Fischer and Ravizza, *Responsibility and Control*, p. 74.

[8] We are grateful to an anonymous referee for raising this worry.

[9] Fischer and Ravizza, *Responsibility and Control*, p. 71, n. 13.

[10] See, for instance, Derk Pereboom, 'Reasons-Responsiveness, Alternative Possibilities, and Manipulation Arguments Against Compatibilism: Reflections on John Martin Fischer's My Way', *Philosophical Books*, 47 (2006), pp. 198-212, Michael McKenna's review of Fischer and Ravizza's book in *Journal of Philosophy*, 98 (2001), pp. 93-100, and Gary Watson, 'Reasons and Responsibility', *Ethics*, 111 (2001), pp. 374-394.

[11] See Fischer, 'The Free Will Revolution (continued)', *The Journal of Ethics*, 10 (2006), pp. 315-345.

[12] We are grateful to Randolph Clarke for this point.

[13] We are grateful to John Martin Fischer for his encouragement and for many helpful discussions about these issues; he was the first to see a version of this worry. We would also like to thank Randolph Clarke, Michael McKenna, and two anonymous referees for providing thoughtful feedback on previous versions of the paper.